

reshape - wide/long 形式変換 【評価版】

`reshape` は wide 形式のデータを long 形式に、long 形式のデータを wide 形式に変換する機能を提供します。

- | | |
|----------------------|------|
| 1. データ形式 | |
| 2. wide から long への変換 | 用例 1 |
| 3. long から wide への変換 | 用例 2 |
| 4. 変則的な用例 | 用例 3 |
| | 用例 4 |

1. データ形式

Stata にはデータの編成に関して **wide 形式**と **long 形式**の区別があります。次の例は売上に関するデータですが、左側が wide 形式による編成、右側が long 形式による編成です。

i			X_{ij}		i		j	X_{ij}
id	company	rev2017	rev2018	rev2019	id	company	year	rev
1	name1	120	130	110	1	name1	2017	120
2	name2	410	320	240	1	name1	2018	130
3	name3	720	770	950	1	name1	2019	110
					2	name2	2017	410
					2	name2	2018	320
					2	name2	2019	240
					3	name3	2017	720
					3	name3	2018	770
					3	name3	2019	950

図 1 wide 形式と long 形式

Stata のコマンドの多くは long 形式の編成を前提としていますが、中には wide 形式を前提としたものもあります。このため両者の間での形式変換の機能が必要となるわけですが、それを提供するのが **reshape** コマンドです。なお、その操作において i 変数、 j 変数、 X_{ij} 変数に対する認識がキーとなるため、図 1 に示された対応関係には十分注意してください。ただし wide 形式の場合、 j 変数は単独で存在するわけではなく少々わかりにくいですが、 X_{ij} の列 j を規定する変数として解釈することができます。

2. wide から long への変換

▷ 用例 1

今、企業の売上げ推移データを含む wide 形式のデータセット `rvwide1.dta` が与えられたとします。

```
. use https://www.math-koubou.jp/stata/data19/rvwide1.dta
. list *1
```

	id	company	rev2017	rev2018	rev2019
1.	1	name1	120	130	110
2.	2	name2	410	320	240
3.	3	name3	720	770	950

これを long 形式に変換するに際しては次の点を認識する必要があります。

- i 変数は `id` である。
- X_{ij} 変数は `rev2017`, `rev2018`, `rev2019` の 3 つであり、その共通名称 (stub) は `rev` である。
- X_{ij} 変数の列方向を識別する j 変数の値は 2017, 2018, 2019 である。このためその変数名としては `year` が適当である。

そこで **reshape** コマンドの実行に際して次のような指定を行います。

- Data ▷ Create or change data ▷ Other variable-transformation commands
 - ▷ Convert data between wide and long と操作
- reshape ダイアログ: Long format from wide (デフォルト)
 - ID variable(s), option `i()`: `id`
 - Subobservation identifier, option `j()`: Variable: `year`
 - Base (stub) names of X_{ij} variables: `rev`

^{*1} メニュー操作: Data ▷ Describe data ▷ List data

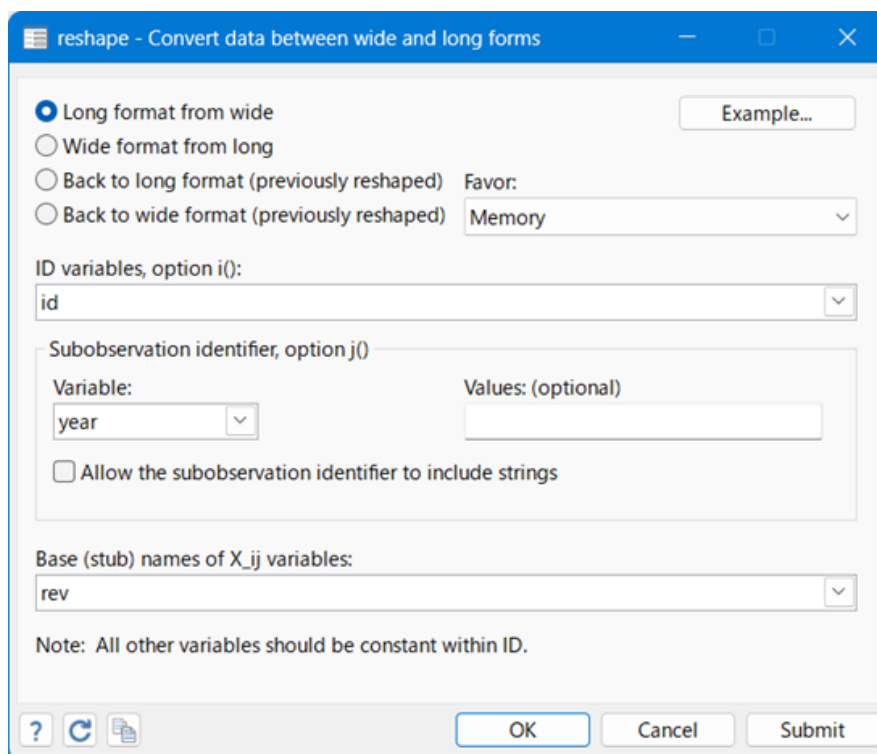


図 2 reshape ダイアログ

```
. reshape long rev, i(id) j(year)
(j = 2017 2018 2019)
```

Data	Wide	->	Long
Number of observations	3	->	9
Number of variables	5	->	4
j variable (3 values)		->	year
xij variables:			
	rev2017 rev2018 rev2019	->	rev

この変換操作によってデータセットの内容は次のように変化します。

```
. list, sepby(id)
```

	id	year	company	rev
1.	1	2017	name1	120
2.	1	2018	name1	130
3.	1	2019	name1	110
4.	2	2017	name2	410
5.	2	2018	name2	320
6.	2	2019	name2	240
7.	3	2017	name3	720
8.	3	2018	name3	770
9.	3	2019	name3	950

正しく long 形式に変換されていることはわかると思いますが、変数 `year` と `rev` が離れているのが気になるかも知れません。変数の並びは `order` コマンドによって調整できます。

- Data ▸ Data utilities ▸ Change order of variables と操作
- `order` ダイアログ: Variables to reorder: year
Place variables to reorder before specified variable: rev

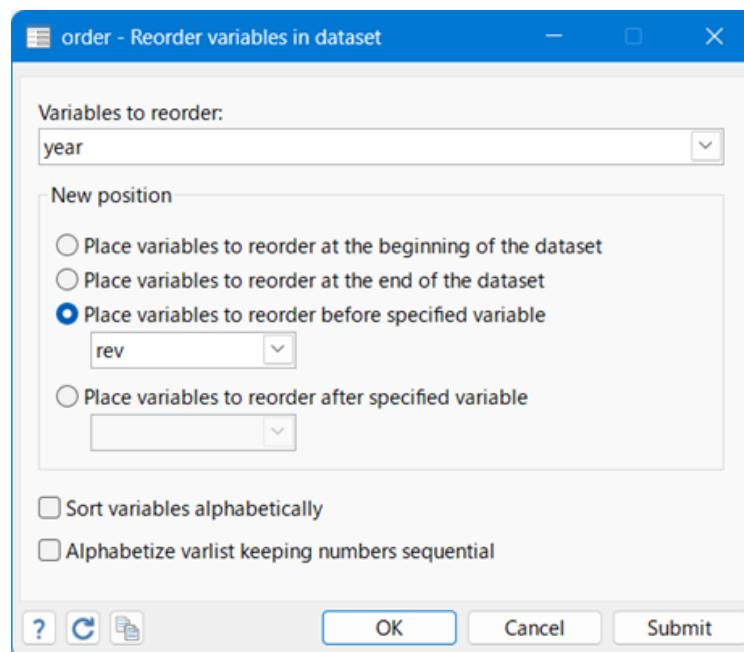


図 3 `order` ダイアログ

```
. order year, before(rev)
```

この状態で再度データセットの内容をリスト出力しておきます。

```
. list, sepby(id)
```

	id	company	year	rev
1.	1	name1	2017	120
2.	1	name1	2018	130
3.	1	name1	2019	110
4.	2	name2	2017	410
5.	2	name2	2018	320
6.	2	name2	2019	240
7.	3	name3	2017	720
8.	3	name3	2018	770
9.	3	name3	2019	950

<



ここでは `reshape long` コマンドを実行したわけですが、その操作を `undo` するには `reshape wide` とコマンド入力します。



この例において `company` という変数は `reshape` の操作には関与していないわけですが、その値は `j` 変数の値によらず一定であるものとみなされます。

3. long から wide への変換

▷ 用例 2

評価版では割愛しています。

4. 変則的な用例

▷ 用例 3

評価版では割愛しています。

▷ 用例 4

評価版では割愛しています。

■